RNA Biology 10:5, 679–686; May 2013; © 2013 Landes Bioscience

CRISPR-CAS

Evolution of an RNA-based adaptive immunity system in prokaryotes

Eugene V. Koonin* and Kira S. Makarova

National Center for Biotechnology Information; NLM; National Institutes of Health; Bethesda, MD USA

Keywords: CRISPR-Cas, adaptive immunity, innate immunity, programmed cell death, dormancy, RRM domain

The CRISPR-Cas (clustered regularly interspaced short palindromic repeats, CRISPR-associated genes) is an adaptive immunity system in bacteria and archaea that functions via a distinct self-non-self recognition mechanism that is partially analogous to the mechanism of eukaryotic RNA interference (RNAi). The CRISPR-Cas system incorporates fragments of virus or plasmid DNA into the CRISPR repeat cassettes and employs the processed transcripts of these spacers as guide RNAs to cleave the cognate foreign DNA or RNA. The Cas proteins, however, are not homologous to the proteins involved in RNAi and comprise numerous, highly diverged families. The majority of the Cas proteins contain diverse variants of the RNA recognition motif (RRM), a widespread RNA-binding domain. Despite the fast evolution that is typical of the *cas* genes, the presence of diverse versions of the RRM in most Cas proteins provides for a simple scenario for the evolution of the three distinct types of CRISPR-cas systems. In addition to several proteins that are directly implicated in the immune response, the *cas* genes encode a variety of proteins that are homologous to prokaryotic toxins that typically possess nuclease activity. The predicted toxins associated with CRISPR-Cas systems include the essential Cas2 protein, proteins of COG1517 that, in addition to a ligand-binding domain and a helix-turn-helix domain, typically contain different nuclease domains and several other predicted nucleases. The tight association of the CRISPR-Cas immunity systems with predicted toxins that, upon activation, would induce dormancy or cell death suggests that adaptive immunity and dormancy/suicide response are functionally coupled. Such coupling could manifest in the persistence state being induced and potentially providing conditions for more effective action of the immune system or in cell death being triggered when immunity fails.

Introduction

The CRISPR-Cas system employs a unique defense mechanism that involves incorporation of virus DNA fragments into CRISPR repeat arrays and subsequent utilization of transcripts of these inserts (spacers) as guide RNAs to cleave the cognate virus genome.[1-5] Thus, CRISPR-Cas represents bona fide adaptive immunity that until the discovery of this system, has not been known to exist in prokaryotes.[6] However, an important distinction between CRISPR-Cas and animal immune systems is that CRISPR-Cas modifies the organism's genome in response to infection and, hence, provides heritable immunity. Accordingly, CRISPR-Cas is the most obvious known case of Lamarckian inheritance whereby an organism reacts to an environmental cue by generating a heritable modification of the genome that provides an adaptive response to that specific cue.[7] The role in antivirus defense that initially was predicted for the CRISPR-Cas system on the basis of the detection of spacers identical to short sequence segments from virus and plasmid genomes and comparative analysis of Cas protein sequences[4] has been successfully demonstrated experimentally.[8] In the few years that elapsed since this key breakthrough, the CRISPR research has evolved into a highly dynamic field of microbiology with major potential for applications in epidemiology, biotechnology and genome engineering.[9-12] The first applications of CRIPSR-Cas for genome manipulation and gene expression programming have already been developed.[13-15]

The CRISPR-Cas systems are classified into three distinct types (I, II and III)[16] and several still unclassified minor variants.[17] This classification was developed through a combination of sequence comparison of the Cas proteins, *cas* gene repertoires and genomic organization of the *CRISPR-cas* loci. For each type and subtype, a specific signature gene has been identified allowing easy classification of the highly variable CRISPR-Cas loci in the course of genome analysis.[16]

The mechanism of CRISPR-Cas is usually divided into three stages: (1) adaptation, when new spacers homologous to protospacer sequences in viral genomes or other alien DNA molecules are integrated into the CRISPR repeat cassettes; (2) expression and processing of pre-crRNA into short guide crRNAs and (3) interference, when the alien DNA or RNA is targeted by a complex containing a crRNA guide and a set of Cas proteins.[16,18,19]

The recent advances in the study of CRISPR-Cas systems are covered in many review articles,[19-23] so here we present only a brief outline of the functions and comparative genomics of prokaryotic adaptive immunity, and discuss the likely scenarios for the evolution of the different types of CRISPR-Cas systems. We then focus on the basic building blocks of the three types of CRISPR-Cas systems and summarize the current understanding of the

*Correspondence to: Eugene V. Koonin; Email: koonin@ncbi.nlm.nih.gov
Submitted: 01/10/13; Revised: 02/12/13; Accepted: 02/15/13
http://dx.doi.org/10.4161/rna.24022

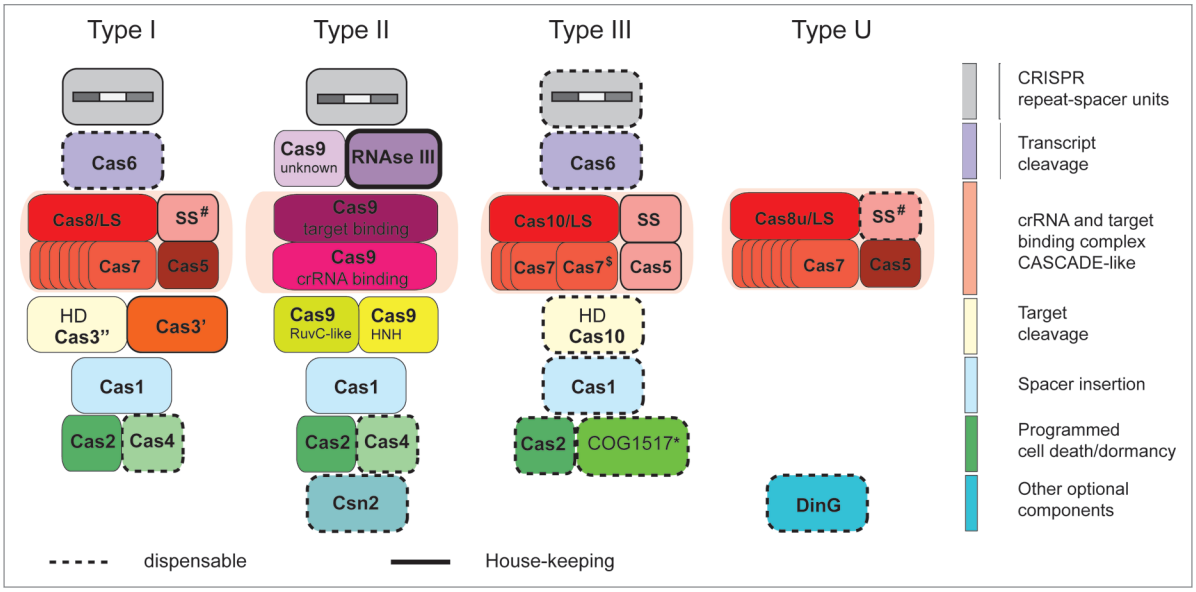

Figure 1. The principal building blocks of the three types of CRISPR-Cas systems. For each Cas protein, the systematic gene name[16] and (whenever available) domain description are given. The color code for general functions is shown on the right. Homologous genes are shown by shapes of the same color, shades denote specialized activities within the same general function. Abbreviations: LS, large sunubit (including Cas10, Cas8, Cse1 and Csy1 subfamilies); SS, small subunit. The # symbol indicates putative small subunits that appear to be fused to the large subunit in several Type I subtypes.[17] The $ symbol indicates that several distinct subfamilies of the Cas7 family could be present in a Cascade complex with unknown stoichiometry. An asterisk indicates that the respective COG1517 family proteins contain a third effector (toxin) domain.

origin and evolution of CRISPR-Cas, and its interaction and coevolution with other types of defense systems in prokaryotes.

Diversity and Hidden Unity Among Cas Protein Domains

A recent comprehensive analysis of the selection factors that affect the evolution *cas* genes has shown that Cas protein sequences evolve under relaxed purifying selection, far below the genomic mean for the respective lineages of bacteria and archaea.[24] This weakness of selective constraints results in fast evolution and accounts for the considerable difficulty of the classification of Cas proteins on the basis of sequence or even structural similarity. Conventional sequence comparison partitions the Cas proteins into more than 100 families. However, the use of more sophisticated methods for sequence analysis combined with structural comparison provided for the detection of many shared domains between Cas protein families that were originally considered unrelated and, thus, enabled the identification of the major building blocks for each type of CRISPR-Cas systems (**Fig. 1**).[1,4,16,17]

The only two proteins that are present in all CRISPR-Cas systems are Cas1 and Cas2 that together are responsible for spacer integration.[25] Although both Cas1 and Cas2 are involved in and are essential for the adaptation stage of the CRISPR response, Cas1 appears to possess all the required enzymatic activities. Specifically, Cas1 is a metal-dependent DNase that produces DNA fragments approximately 80 base pairs in size.[26] The unique protein fold of the Cas1 endonuclease is the most highly conserved domain among the Cas proteins. Accordingly, Cas1 is the best phylogenetic marker for the study of the evolution of CRISPR-Cas systems. Generally, Cas1 phylogeny correlates well with CRISPR-Cas subtype classification; only *cas1* genes associated with type III systems are polyphyletic.[16] Together with the fact that type III systems often accompany Type I system within the same genome, this observation implies that Type III systems might utilize the CRISPR arrays generated by type I systems and, accordingly, the same Cas1 protein could function with both CRISPR-Cas types. The only CRISPR-Cas system that encompasses neither Cas1 nor a CRISPR array is Type U for which the function and the mechanism of action remain unclear.[16,17]

With the exception of Cas1, most of the common Cas proteins contain diverse versions of the RNA recognition motif (RRM) domain, a RNA-binding domain that is widely represented in all cellular life forms and in particular comprises the core of diverse DNA and RNA polymerases (where it is known as the Palm domain). Among the Cas proteins, different variants of the RRM domain are present in Cas2 (a toxin-like ribonuclease),[17,27] Cas10 (the so-called CRISPR polymerase, a protein that is homologous to polymerases and cyclases[17,28-30] but whose actual biochemical activity remain unknown) and in the largest group of Cas proteins known as the RAMP (repeat associated mysterious proteins) superfamily (**Fig. 2**). In all Type I CRISPR-Cas systems and in most of the Type III systems, there is a dedicated ribonuclease for the pre-crRNA processing, typically a protein of the Cas6 family of the RAMPs. In some cases, e.g., in the Type I-C CRISPR-Cas systems, Cas6 is displaced by another catalytically active RAMP, in particular Cas5.[31-33] Type II CRISPR-Cas systems employ an unrelated mechanism of pre-crRNA cleavage that requires the activity

of the double-stranded RNA-specific RNase III, a specialized trans-encoded small RNA (tracrRNA), which is complementary to a single CRISPR unit, and still unidentified domains of the large Cas9 protein.[17,34-36]

In Type I CRISPR-Cas systems (at least Type I-E and Type I-F), the endoribonuclease that catalyzes the processing of the pre-crRNA is a subunit of a multisubunit (or multidomain) complex known as CASCADE (CRISPR-associated complex for antiviral defense).[37,38] The mature crRNA remains associated with the CASCADE complex that scans the target DNA for a match, and once one is found, cleaves the target.[21] In Type III systems (at least the model system from the archaeon *Pyrococcus furiosus*), the Cas6 endoribonuclease does not belong to the CASCADE complex and functions as an independent protein.[39,40] Type III CASCADE apparently is not directly involved in the processing but instead binds the mature crRNA and cleaves alien RNA.[41] Moreover, in those archaea that encode Type III along with Type I-A or I-B CRISPR-Cas systems, Cas6 might be shared between the two types.[40] These distinctions apart, the subunit architectures of the CASCADE complexes in Type I and Type III CRISPR-Cas systems are similar and include a large subunit, a small subunit and a pair of RAMPs that belong to the Cas5 and Cas7 families[33,39,42-44] (**Fig. 1**). However, the protein subunits themselves are highly diverged.[4,16,17]

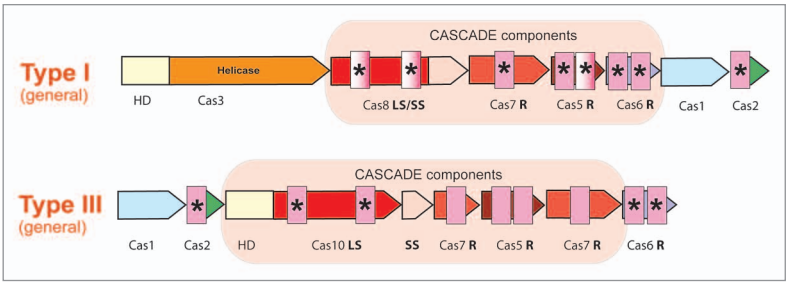

Figure 2. The RRM domain, the cornerstone of CRISPR-Cas. The figure shows RRM domain-containing proteins in CRISPR-Cas systems. The typical organization of the operons in Type I and Type III CRISPR-Cas loci is depicted by arrows with the size roughly proportional to the size of respective gene. Homologous genes are shown by arrows of same color. Color coding is the same as in the **Figure 1**. Gene and family names are from references 16 and 17. Additional designations: LS, large subunit; SS, small subunit; R, RAMPs. RRM domains are shown by pink rectangles, with semi-transparent rectangles indicating deteriorated RRM fold. The proteins representing families with RRM domains for which structures have been solved are denoted by asterisks.

Despite the high level of sequence divergence that is typical of most Cas proteins, there appears to be a direct homologous relationship between the respective subunits of the Type I and Type III CASCADEs.[17] The most complicated relationship involves the large subunits that show no readily detectable sequence or structural similarity between Type I and Type III CRISPR-Cas systems.[28,29,45] Nevertheless, the similarity in the overall architectures of these proteins (KSM, unpublished) suggests that the large subunit of Type I systems (Cse1) might be an inactivated, highly derived form of the Type III large subunit (Cas10) that is predicted to be an active cyclase/polymerase-like enzyme.[4,28,29] Another notable difference is that Type I CRISPR-Cas encompasses a single Cas7 protein that is represented by multiple copies in the CASCADE, whereas in Type III systems there are several paralogous Cas7-like proteins (**Fig. 1**).

In Type II CRISPR-Cas, a single large, multidomain protein, Cas9, is responsible for all the functions of the CASCADE complex subunits as well as the function of Cas6 (when distinct from CASCADE).[34,46] The fusion of all domains that are required for the target cleavage within a single protein makes Type II CRISPR-Cas the system of choice for application in genome manipulation.[14,15]

The target DNA cleavage in Type I,[47] and most likely in Type III systems,[17] is catalyzed by homologous HD family nucleases. In many Type III systems, the HD domain is fused to the Cas10 gene, the large subunit of the CASCADE-like complex, whereas in Type I systems, the most common protein architecture involved in cleavage is Cas3 in which the HD domain is fused to a distinct helicase domain that is essential for the interference stage.[47,48] Type II systems employ an unrelated mechanism that involves two distinct nuclease domains, HNH and RuvC-like, both contained within the Cas9 protein.[49] The recognition of the target non-self DNA is apparently mediated by the large subunit of the Type I Cascade complex that recognizes the so-called PAM (protospacer-associated motif) and promotes loop formation in the target DNA.[45]

The Cas1 endonuclease, the CASCADE complex subunits and the Cas3 helicase-nuclease are essential for the immune function of the respective CRISPR-Cas systems. In addition, the CRISPR-Cas loci encompass a remarkable variety of genes that encode proteins whose mechanistic role in CRISPR-dependent adaptive immunity remains unclear but that belong to protein families implicated in other defense systems. These CRISPR-associated genes include the ribonuclease Cas2, the RecB-like nuclease Cas4, for which a 5' to 3'DNA exonuclease activity has been demonstrated[50] and numerous representatives of the COG1517 superfamily proteins that contain a helix-turn-helix domain often combined with various enzymatic domains.[4,17] Most of these proteins, in particular Cas2, contain domains that are predicted to be nucleases and toxins, suggesting a secondary role as associated immunity components (see details below and ref. 51). The single-stranded RNA-specific ribonuclease activity of Cas2 from a Type I system of the archaeon *Sulfolobus solfataricus* has been demonstrated experimentally,[27] and similar results have been subsequently obtained for the bacterium *Helicobacter pylori*.[52] However, for the Cas2 protein from the Type I-C system of the bacterium *Bacillus halodurans* specificity toward dsDNA has been reported.[33] In contrast, Cas2 of the bacterium *Desulfovibrio vulgaris* lacks enzymatic activity and encompasses replacements of some of the essential catalytic residues;[53] similar, supposedly inactivated forms of Cas2 were detected in several other bacteria.[4] These findings emphasize the extreme functional diversity of CRISPR-Cas systems and might be related to the observations that Type I systems seem to target exclusively

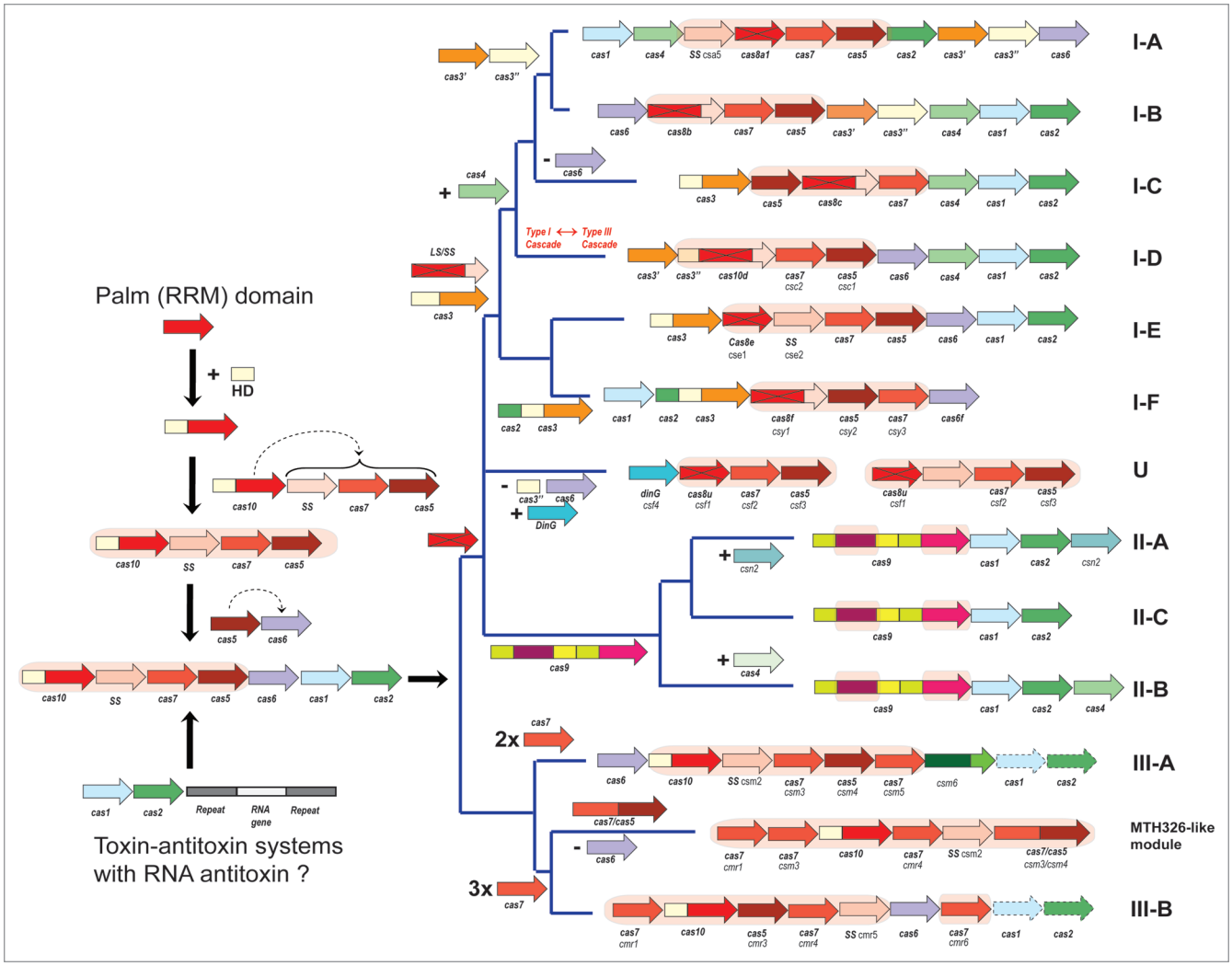

Figure 3. A parsimonious evolutionary scenario for the three types of CRISPR-Cas. Homologous genes are color-coded and identified by a family name.[16,17] Color coding is the same as in **Figure 1**. Genes coding for inactivated (putative) large subunits are indicated by crosses. Major evolutionary events are shown in the corresponding branches. The figure is an updated and modified version of figure 7 in reference 17.

DNA, whereas for Type III systems, RNA targeting has been demonstrated.[41,44]

Finally, the functions of several Cas proteins remain completely obscure. One of these is the inactivated ATPase homolog Csn2, which binds dsDNA as a tetrameric ring; this protein is an optional component of Type II systems[54,55] but its function in CRISPR-mediated immunity remains enigmatic.

Taken together, the results of comparative sequence analysis, structural studies and experimental data suggest that despite the remarkable complexity and diversity, all Type I and Type III CRISPR-Cas systems employ the same architectural and functional principles, and given the conservation of the principal building blocks (**Figs. 1 and 2**), share a common ancestry. It is notable, however, that some of the essential components of the CRISPR-Cas systems can be replaced either by homologous proteins, such as the substitution of Cas5 for Cas6 in Type 1C CASCADE complexes. The most dramatic transformation of the Cas protein machinery is evident in Type II systems in which the CASCADE complex and the effector HD nuclease are replaced with non-homologous but functionally analogous domains of the Cas9 protein, in particular the HNH and RuvC-like nuclease domains.

In the recently derived evolutionary scenario (**Fig. 3**), only a few evolutionary events suffice to explain the emergence of CRISPR-Cas system types and subtypes.[17] Comparison of the recently solved structures of all major components of the CASCADE complexes suggests that the RAMPs and the small subunits might have evolved, respectively, from the ancestral large subunit resembling the Cas10 polymerase-like protein that contains two RRM domains and an alpha-helical domain resembling the small subunit.[28,29] The Cas10 protein could have evolved from an ancestor RRM (Palm) domain-containing polymerase or cyclase, and combined with the HD domain, might have originally functioned as a CRISPR-independent defense system that would represent innate immunity rather than adaptive immunity.[17]

The *cas1-cas2* gene pair strikingly resembles the toxin-antitoxin (TA) two-gene modules in that the two genes are almost

always adjacent and Cas2 encodes a homolog of toxin mRNA interferases.[56-58] It seems a distinct possibility that the Cas1-Cas2 module originally functioned as an independent TA system. Moreover, the CRISPR repeats themselves might have originated from an RNA molecule that originally antitoxin served as an antitoxin to the Cas2 toxin analogously to Type III TA systems.[59-61] Joining this module with the hypothetical ancestral CASCADE-HD system might have led to the emergence of the adaptation stage and, accordingly, the transformation of an innate immunity mechanism into a mechanism of adaptive immunity (see discussion below).

The Cas9 protein is homologous to a large family of prokaryotic proteins that also contain both RuvC-like and HNH-nuclease domain.[17] Most likely, by analogy with Cas9, these proteins are DNases that are involved in an unknown, CRISPR-independent innate immunity mechanism.

The ancestral Cas10-like protein and the entire subtype III-like CRISPR-Cas system most likely evolved in hyperthermophilic archaea and was subsequently horizontally transferred to bacteria. Indeed, in archaeal hyperthermophiles, this variant of the CRISPR-Cas system is (nearly) universal, in a sharp contrast to the presence of CRISPR-Cas in less than 50% of archaeal and bacterial mesophiles.[16,17,62] In accord with this scenario, a recent mathematical modeling study has shown that the benefits of adaptive immunity are substantially greater under the conditions of limited virus variability that seems to be characteristic of hyperthermophilic habitats.[63]

Coupling Between Immunity and Dormancy/Suicide Defense Systems in Prokaryotes

All archaeal and bacterial defense systems can be classified into two major types that function on distinct principles: (1) immune systems that discriminate self DNA from non-self DNA and specifically destroy the non-self, in particular viral genomes, whereas the host genome is protected, (2) systems that induce dormancy or programmed cell suicide in response to infection (and other forms of stress), known as toxin-antitoxin or abortive infection systems.[56,57,64] Most of the genomic loci that encode adaptive immunity systems, namely all types of CRISPR-Cas, and innate immunity systems, such as restriction-modification or DNA phosphothiolation, also encompass genes that encode toxins, in particular nucleases implicated in the induction of dormancy or cell death, mostly via inactivation of the translation system.[51] The most common among these immunity-associated (putative) toxins are proteins containing the HEPN (higher eukaryotes and prokaryotes nucleotide-binding[65]) domain (**Fig. 4**). Detailed comparative sequence and structure analysis leads to a strong prediction that the majority of the HEPN domain possesses RNase activity.[66] However, the immunity loci do not seem to encode antitoxins, at least not those from well-characterized antitoxin families. There is no indication that the toxins are mechanistically involved in the immune functions suggesting that these immunity-associated toxins might perform their typical function in coordination with the respective immunity systems.[51]

According to the immunity-dormancy/suicide coupling hypothesis, the toxins associated with immune systems induce dormancy or cell suicide unless controlled by components of the respective immunity system that act as antitoxins[51] (**Fig. 4**). An experimentally characterized case of such coupling is the interaction between the *Escherichia coli* anticodon nuclease PrrC and PrrI restriction-modification system.[67,68] The coupling of diverse immunity and dormancy/suicide systems could have evolved under selective pressure to provide robustness to the antivirus response. The involvement of dormancy/suicide systems in the coupled antivirus response might take two distinct forms: (1) induction of a dormancy-like state in the infected cell to "buy time" for the activation of adaptive immunity; (2) "altruistic" induction of dormancy or suicide as the final recourse to prevent viral spread triggered by the failure of immunity.

The function of Cas2 would follow the first route, given that Cas2 is essential for the adaptation stage of adaptive immunity[25] and homologous to toxin interferases.[4] Conceivably, one of the Cas proteins, possibly Cas1, acts as an antitoxin to Cas2, but the nuclease activity of Cas2 is switched on when the CRISPR-Cas system encounters a new virus so that the Cas1 protein has an opportunity to detect and insert a new spacer. The dormancy-like response mediated by the activity of Cas2 and/or a COG1517 protein containing an effector domain, of which the most common are the HEPN, RelE and the PD-(D/E)xK (RecB-like) family nucleases (ref. 51 and KSM, unpublished), would prevent virus reproduction allowing the host the time required to prime the immunity response, which could be a relatively slow and inefficient process.

Under the second coupling mode, when an immunity system fails and/or the level of genotoxic stress increases, the cell employs the associated toxins for abrogation of key cell processes, typically translation, resulting in persistence or cell death. The cell suicide in such a case can be considered altruistic, i.e., preventing infection of other bacterial or archaeal cells within the same colony or community. Conceivably, some of the toxin-like proteins encoded within the CRISPR-Cas loci function via this route, and it cannot be ruled out that the toxin activity of Cas2 plays a dual role in adaptive immunity. The functional relevance of the reported DNase activity of Cas2 from *B. halodurans*[33] in the context of this scheme remains to be determined. Furthermore, some versions of Cas2 in Type I and Type II systems are predicted to be inactivated nucleases,[4] possibly suggestive of an alternative mechanism of toxin-like activity.

Remarkably, a plasmid-encoded CRISPR-Cas system containing a RelE toxin in place of Cas2 has been recently identified in *Lactococcus lactis*[69] (**Fig. 4**). This finding reinforces the link between CRISPR-Cas systems and TA systems and is compatible with the hypothesis of adaptive immunity and programmed-cell death/dormancy mechanisms coupling.[51]

Concluding Remarks

The study of the CRISPR-Cas response has been one of the most dynamic areas of microbiology in the past 5–6 y. The CRISPR-Cas systems embody two fundamental biological phenomena,

Figure 4. Hypothetical coupling of the CRISPR-mediated adaptive immunity with the persistence/cell suicide response. The genes are depicted by black arrows. Persistence/cell suicide-related genes are shown in green shades. Homologous genes are shown by the arrows of same color. The PD-(D/E)xK (RecB-like) endonuclease, VapD/Cas2 endoribonuclease and predicted endoribonuclease HEPN domains[51,66] are indicated above the corresponding gene.

adaptive immunity and Lamarckian inheritance, the first of which has not been previously recognized in prokaryotes, whereas the second had been generally considered non-existent. In retrospect, several additional phenomena, in particular, at least some forms of horizontal gene transfer, can be recognized as "quasi-Lamarckian" but CRISPR-Cas arguably remains the most clear-cut case for this mode of evolution.[7] Notably, however, the CRISPR-Cas response also might involve a substantial selective component whereby numerous spacers are initially incorporated in response to virus infection but only a few are fixed in the host population.[70]

Although many key molecular details of the CRISPR-Cas mechanism have been deciphered, in particular, the structure and function of the CASCADE complexes, other important aspects of the CRISPR-Cas function remain to be understood, especially with respect to self-non-self discrimination. The *cas* genes are subject to weak selection pressure and, accordingly, are among the fastest evolving genes in prokaryote genomes. Accordingly, most of the Cas proteins show limited sequence conservation so that sophisticated sequence and structure comparison methods are essential to establish homologous relationships. Nevertheless, the presence of diverse versions of the RRM domain in most of

the Cas proteins provides for unification of these proteins into families and superfamilies and for a relatively simple evolutionary scenario that accounts for the origin of the three types of CRISPR-Cas systems from a stand-alone, CRISPR-independent, CASCADE-like innate immunity system and a toxin-antitoxin-like Cas1-Cas2 module.

In addition to several proteins that are directly implicated in the adaptive immunity response, the *cas* genes encode a variety of proteins that are homologous to prokaryotic toxins, most of which possess nuclease activity. The predicted toxins associated with CRISPR-Cas systems include the essential Cas2 protein and several other predicted nucleases. The tight association of the CRISPR-Cas immunity systems with predicted toxins that, upon activation, would induce dormancy or cell death implies functional coupling between adaptive immunity and dormancy/suicide response. Such coupling could involve induction of a persistence-like state that could provide for more effective action of the immune system or in induction of cell death when immunity fails. This immunity-dormancy/suicide coupling hypothesis implies many experimentally testable predictions on the biological functions of Cas proteins.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors' research is supported by the intramural program of the United States Department of Health and Human Services (to National Library of Medicine).

References

1. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol 2005; 1:e60; PMID:16292354; http://dx.doi.org/10.1371/journal.pcbi.0010060.
2. Jansen R, Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 2002; 43:1565-75; PMID:11952905; http://dx.doi.org/10.1046/j.1365-2958.2002.02839.x.
3. Koonin EV, Makarova KS. CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 Biol Rep* 2009; 1:95.
4. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct 2006; 1:7; PMID:16545108; http://dx.doi.org/10.1186/1745-6150-1-7.
5. Mojica FJ, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol 2005; 60:174-82; PMID:15791728; http://dx.doi.org/10.1007/s00239-004-0046-3.
6. Goren M, Yosef I, Edgar R, Qimron U. The bacterial CRISPR/Cas system as analog of the mammalian adaptive immune system. RNA Biol 2012; 9:549-54; PMID:22614830; http://dx.doi.org/10.4161/rna.20177.
7. Koonin EV, Wolf YI. Is evolution Darwinian or/and Lamarckian? Biol Direct 2009; 4:42; PMID:19906303; http://dx.doi.org/10.1186/1745-6150-4-42.
8. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007; 315:1709-12; PMID:17379808; http://dx.doi.org/10.1126/science.1138140.
9. Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, et al. CRISPR typing and subtyping for improved laboratory surveillance of Salmonella infections. PLoS One 2012; 7:e36995; PMID:22623967; http://dx.doi.org/10.1371/journal.pone.0036995.
10. Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. Cell Host Microbe 2012; 12:177-86; PMID:22901538; http://dx.doi.org/10.1016/j.chom.2012.06.003.
11. Weinberger AD, Gilmore MS. CRISPR-Cas: to take up DNA or not-that is the question. Cell Host Microbe 2012; 12:125-6; PMID:22901532; http://dx.doi.org/10.1016/j.chom.2012.07.007.
12. Carroll D. A CRISPR approach to gene targeting. Mol Ther 1658-1660; 2012:mt2012171; PMID:22945229.
13. Qi L, Haurwitz RE, Shao W, Doudna JA, Arkin AP. RNA processing enables predictable programming of gene expression. Nat Biotechnol 2012; 30:1002-6; PMID:22983090; http://dx.doi.org/10.1038/nbt.2355.
14. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science 2013; 339:823-6; PMID:23287722; http://dx.doi.org/10.1126/science.1232033.
15. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science 2013; 339:819-23; PMID:23287718; http://dx.doi.org/10.1126/science.1231143.
16. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 2011; 9:467-77; PMID:21552286; http://dx.doi.org/10.1038/nrmicro2577.
17. Makarova KS, Aravind L, Wolf YI, Koonin EV. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol Direct 2011; 6:38; PMID:21756346; http://dx.doi.org/10.1186/1745-6150-6-38.
18. van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci 2009; 34:401-7; PMID:19646880; http://dx.doi.org/10.1016/j.tibs.2009.05.002.
19. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature 2012; 482:331-8; PMID:22337052; http://dx.doi.org/10.1038/nature10886.
20. Fineran PC, Charpentier E. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. Virology 2012; 434:202-9; PMID:23123013; http://dx.doi.org/10.1016/j.virol.2012.10.003.
21. Westra ER, van Erp PB, Künne T, Wong SP, Staals RH, Seegers CL, et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell 2012; 46:595-605; PMID:22521689; http://dx.doi.org/10.1016/j.molcel.2012.03.018.
22. Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 2011; 45:273-97; PMID:22060043; http://dx.doi.org/10.1146/annurev-genet-110410-132430.
23. Barrangou R, Van der Oost J, eds. CRISPR-Cas Systems. RNA-mediated Adaptive Immunity in Bacteria and Archaea. (Springer, Heidelberg, 2013).
24. Takeuchi N, Wolf YI, Makarova KS, Koonin EV. Nature and intensity of selection pressure on CRISPR-associated genes. J Bacteriol 2012; 194:1216-25; PMID:22178975; http://dx.doi.org/10.1128/JB.06521-11.
25. Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res 2012; 40:5569-76; PMID:22402487; http://dx.doi.org/10.1093/nar/gks216.
26. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 2009; 17:904-12; PMID:19523907; http://dx.doi.org/10.1016/j.str.2009.03.019.
27. Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, et al. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J Biol Chem 2008; 283:20361-71; PMID:18482976; http://dx.doi.org/10.1074/jbc.M803225200.
28. Zhu X, Ye K. Crystal structure of Cmr2 suggests a nucleotide cyclase-related enzyme in type III CRISPR-Cas systems. FEBS Lett 2012; 586:939-45; PMID:22449983; http://dx.doi.org/10.1016/j.febslet.2012.02.036.
29. Cocozaki AI, Ramia NF, Shao Y, Hale CR, Terns RM, Terns MP, et al. Structure of the Cmr2 subunit of the CRISPR-Cas RNA silencing complex. Structure 2012; 20:545-53; PMID:22405013; http://dx.doi.org/10.1016/j.str.2012.01.018.
30. Huang RH. Cas protein Cmr2 full of surprises. Structure 2012; 20:389-90; PMID:22404997; http://dx.doi.org/10.1016/j.str.2012.02.007.
31. Carte J, Pfister NT, Compton MM, Terns RM, Terns MP. Binding and cleavage of CRISPR RNA by Cas6. RNA 2010; 16:2181-8; PMID:20884784; http://dx.doi.org/10.1261/rna.2230110.
32. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science 2010; 329:1355-8; PMID:20829488; http://dx.doi.org/10.1126/science.1192272.
33. Nam KH, Ding F, Haitjema C, Huang Q, DeLisa MP, Ke A. Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. J Biol Chem 2012; 287:35943-52.

34. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 2011; 471:602-7; PMID:21455174; http://dx.doi.org/10.1038/nature09886.
35. Gottesman S. Microbiology: Dicing defence in bacteria. Nature 2011; 471:588-9; PMID:21455171; http://dx.doi.org/10.1038/471588a.
36. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 2010; 468:67-71; PMID:21048762; http://dx.doi.org/10.1038/nature09523.
37. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 2008; 321:960-4; PMID:18703739; http://dx.doi.org/10.1126/science.1159689.
38. IvanđčBaĉ I, Howard JA, Bolt EL. Tuning in to interference: R-loops and cascade complexes in CRISPR immunity. J Mol Biol 2012; 422:607-16; PMID:22743103; http://dx.doi.org/10.1016/j.jmb.2012.06.024.
39. Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, et al. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Mol Cell 2012; 45:303-13; PMID:22227115; http://dx.doi.org/10.1016/j.molcel.2011.12.013.
40. Deng L, Garrett RA, Shah SA, Peng X, She Q. A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. Mol Microbiol 2013; In press; PMID:23320564; http://dx.doi.org/10.1111/mmi.12152.
41. Wang R, Preamplume G, Terns MP, Terns RM, Li H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. Structure 2011; 19:257-64; PMID:21300293; http://dx.doi.org/10.1016/j.str.2010.11.014.
42. van Duijn E, Barbu IM, Barendregt A, Jore MM, Wiedenheft B, Lundgren M, et al. Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated protein complexes from Escherichia coli and Pseudomonas aeruginosa. Mol Cell Proteomics 2012; 11:1430-41; PMID:22918228; http://dx.doi.org/10.1074/mcp.M112.020263.
43. Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 2011; 477:486-9; PMID:21938068; http://dx.doi.org/10.1038/nature10402.
44. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 2009; 139:945-56; PMID:19945378; http://dx.doi.org/10.1016/j.cell.2009.07.040.
45. Sashital DG, Wiedenheft B, Doudna JA. Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol Cell 2012; 46:606-15; PMID:22521690; http://dx.doi.org/10.1016/j.molcel.2012.03.020.
46. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA 2012; 109:E2579-86; PMID:22949671; http://dx.doi.org/10.1073/pnas.1208507109.
47. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J 2011; 30:1335-42; PMID:21343909; http://dx.doi.org/10.1038/emboj.2011.41.
48. Mulepati S, Bailey S. Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). J Biol Chem 2011; 286:31896-903; PMID:21775431; http://dx.doi.org/10.1074/jbc.M111.270017.
49. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012; 337:816-21; PMID:22745249; http://dx.doi.org/10.1126/science.1225829.
50. Zhang J, Kasciukovic T, White MF. The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. PLoS One 2012; 7:e47232; PMID:23056615; http://dx.doi.org/10.1371/journal.pone.0047232.
51. Makarova KS, Anantharaman V, Aravind L, Koonin EV. Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. Biol Direct 2012; 7:40; PMID:23151069; http://dx.doi.org/10.1186/1745-6150-7-40.
52. Kwon AR, Kim JH, Park SJ, Lee KY, Min YH, Im H, et al. Structural and biochemical characterization of HP0315 from Helicobacter pylori as a VapD protein with an endoribonuclease activity. Nucleic Acids Res 2012; 40:4216-28; PMID:22241770; http://dx.doi.org/10.1093/nar/gkr1305.
53. Samai P, Smith P, Shuman S. Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. Acta Crystallogr Sect F Struct Biol Cryst Commun 2010; 66:1552-6; PMID:21139194; http://dx.doi.org/10.1107/S1744309110039801.
54. Nam KH, Kurinov I, Ke A. Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. J Biol Chem 2011; 286:30759-68; PMID:21697083; http://dx.doi.org/10.1074/jbc.M111.256263.
55. Lee KH, Lee SG, Eun Lee K, Jeon H, Robinson H, Oh BH. Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. Proteins 2012; 80:2573-82; PMID:22753072; http://dx.doi.org/10.1002/prot.24138.
56. Makarova KS, Wolf YI, Koonin EV. Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. Biol Direct 2009; 4:19; PMID:19493340; http://dx.doi.org/10.1186/1745-6150-4-19.
57. Van Melderen L, Saavedra De Bast M. Bacterial toxin-antitoxin systems: more than selfish entities? PLoS Genet 2009; 5:e1000437; PMID:19325885; http://dx.doi.org/10.1371/journal.pgen.1000437.
58. Yamaguchi Y, Inouye M. mRNA interferases, sequence-specific endoribonucleases from the toxin-antitoxin systems. Prog Mol Biol Transl Sci 2009; 85:467-500; PMID:19215780; http://dx.doi.org/10.1016/S0079-6603(08)00812-X.
59. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GP. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci USA 2009; 106:894-9; PMID:19124776; http://dx.doi.org/10.1073/pnas.0808832106.
60. Blower TR, Pei XY, Short FL, Fineran PC, Humphreys DP, Luisi BF, et al. A processed noncoding RNA regulates an altruistic bacterial antiviral system. Nat Struct Mol Biol 2011; 18:185-90; PMID:21240270; http://dx.doi.org/10.1038/nsmb.1981.
61. Short FL, Pei XY, Blower TR, Ong SL, Fineran PC, Luisi BF, et al. Selectivity and self-assembly in the control of a bacterial toxin by an antitoxic noncoding RNA pseudoknot. Proc Natl Acad Sci USA 2012; PMID:23267117.
62. Makarova KS, Wolf YI, Koonin EV. Potential genomic determinants of hyperthermophily. Trends Genet 2003; 19:172-6; PMID:12683966; http://dx.doi.org/10.1016/S0168-9525(03)00047-7.
63. Weinberger AD, Wolf YI, Lobkovsky AE, Gilmore MS, Koonin EV. Viral diversity threshold for adaptive immunity in prokaryotes. MBio 2012; 3:e00456-12; PMID:23221803; http://dx.doi.org/10.1128/mBio.00456-12.
64. Gerdes K, ed. Prokaryotic Toxin-Antitoxins. (Springer, Heidelberg, 2012).
65. Grynberg M, Erlandsen H, Godzik A. HEPN: a common domain in bacterial drug resistance and human neurodegenerative proteins. Trends Biochem Sci 2003; 28:224-6; PMID:12765831; http://dx.doi.org/10.1016/S0968-0004(03)00060-4.
66. Anantharaman V, Makarova KS, Burroughs AM, Koonin EV., Aravind L. HEPN: A major nucleic-acid targeting domain involved in intra-genomic conflicts, defense, pathogenesis and RNA processing. Biol Direct 2013; In press.
67. Kaufmann G. Anticodon nucleases. Trends Biochem Sci 2000; 25:70-4; PMID:10664586; http://dx.doi.org/10.1016/S0968-0004(99)01525-X.
68. Uzan M, Miller ES. Post-transcriptional control by bacteriophage T4: mRNA decay and inhibition of translation initiation. Virol J 2010; 7:360; PMID:21129205; http://dx.doi.org/10.1186/1743-422X-7-360.
69. Millen AM, Horvath P, Boyaval P, Romero DA. Mobile CRISPR/Cas-mediated bacteriophage resistance in Lactococcus lactis. PLoS One 2012; 7:e51663; PMID:23240053; http://dx.doi.org/10.1371/journal.pone.0051663.
70. Paez-Espino D, Morovic W, Sun CL, Thomas BC, Ueda K, Stahl B, et al. Strong bias in the bacterial CRISPR elements that confer immunity to phage. Nat Commun 2013; 4:1430; PMID:23385575; http://dx.doi.org/10.1038/ncomms2440.